

## Article

# Understanding FBA Solutions under Multiple Nutrient Limitations

Eunice van Pelt-KleinJan <sup>1,2,†</sup> , Daan H. de Groot <sup>2,†,‡</sup>  and Bas Teusink <sup>1,2,\*</sup> <sup>1</sup> TiFN, P.O. Box 557, NL6700AN Wageningen, The Netherlands; e.van.pelt-kleinjan@vu.nl<sup>2</sup> Systems Biology Lab, Amsterdam Institute of Molecular and Life Sciences, Vrije Universiteit Amsterdam, 1081HZ Amsterdam, The Netherlands; daan.degroot@unibas.ch

\* Correspondence: b.teusink@vu.nl

† These authors contributed equally to this work.

‡ Current address: Biozentrum, University of Basel, 4056 Basel, Switzerland.

**Abstract:** Genome-scale stoichiometric modeling methods, in particular Flux Balance Analysis (FBA) and variations thereof, are widely used to investigate cell metabolism and to optimize biotechnological processes. Given (1) a metabolic network, which can be reconstructed from an organism's genome sequence, and (2) constraints on reaction rates, which may be based on measured nutrient uptake rates, FBA predicts which reactions maximize an objective flux, usually the production of cell components. Although FBA solutions may accurately predict the metabolic behavior of a cell, the actual flux predictions are often hard to interpret. This is especially the case for conditions with many constraints, such as for organisms growing in rich nutrient environments: it remains unclear why a certain solution was optimal. Here, we rationalize FBA solutions by explaining for which properties the optimal combination of metabolic strategies is selected. We provide a graphical formalism in which the selection of solutions can be visualized; we illustrate how this perspective provides a glimpse of the logic that underlies genome-scale modeling by applying our formalism to models of various sizes.

**Keywords:** flux balance analysis; elementary flux modes; elementary conversion modes; genome-scale modeling; stoichiometric modeling; phenotype phase plane analysis



**Citation:** van Pelt-KleinJan, E.; de Groot, D.H.; Teusink, B. Understanding FBA Solutions under Multiple Nutrient Limitations. *Metabolites* **2021**, *11*, 257. <https://doi.org/10.3390/metabo11050257>

Academic Editors: Dario Pescini and Marzia Di Filippo

Received: 16 March 2021

Accepted: 19 April 2021

Published: 21 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Flux Balance Analysis (FBA) is a constraint-based optimization method that is used to study metabolic networks in steady state [1,2]. In the most basic FBA, the rate of biomass synthesis is maximized to model metabolic behavior under selection for fast growth. This biomass synthesis is limited by a constraint on some nutrient uptake rate, mimicking the limited availability of this nutrient.

However, Flux Balance Analyses often become much more complicated. For example, because the studied cell can use many different nutrients, in which case there may be many uptake rates with an upper bound, or the modeler may be interested in inferring the full metabolic behavior from the measurement of some reaction rates, which would lead to equality constraints that fix these experimentally measured rates. Alternatively, models may include a growth-independent maintenance parameter via a lower bound on the production of the energy-equivalent ATP. All these cases lead to Flux Balance Analyses with multiple constraints, and a small sample of published models indeed shows that multi-constraint models are abundant (Table 1). When FBA is used with several such constraints, it can become extremely challenging to understand why a certain solution is optimal. This severely obstructs a deep understanding of the model results, and with that, the understanding of the cell's metabolism. It is therefore important to lay bare the logic that underlies FBA solutions. This is what we intend to do in this paper.

A comprehensive list of minimal reaction sets that can sustain a steady-state flux through the network is given by the Elementary Flux Modes (EFMs) [3]. EFMs can be viewed as the minimal metabolic strategies that a cell has at its disposal, and each metabolic state is a combination of these EFMs. Therefore, understanding FBA solutions comes down to rationalizing why the selected EFMs are used. For example, if only one flux constraint is used, often the uptake rate of the sole energy or carbon source, FBA will always select the EFM with the highest yield, i.e., the pathway that produces the most moles of biomass per mole of the limiting substrate [4,5]. Therefore, although FBA appears to explicitly maximize the biomass production rate as a proxy for maximizing the growth rate, the resultant high-yield strategy does not necessarily lead to the fastest biomass production in real organisms [4,6–8], because they may be limited by thermodynamic and enzyme-kinetic limitations [9,10]. Indeed, many microorganisms, and even cancer cells, do not use the high-yield pathways in fast-growth conditions, but rather use seemingly wasteful pathways that, for example, lead to ‘overflow metabolism’ [11–15].

Although it has been well-understood that FBA with one constraint selects the maximal-yield EFM, this logic underlying the solution cannot be readily extended to cases with multiple constraints. We will use small toy models to prove that the highest-yield pathways are not always selected in such cases. Instead, we will show that FBA selects Elementary Flux Modes based on a weighted combination of their product yields for the various constrained uptake rates.

**Table 1.** Published genome-scale models often have several active constraints on reaction rates. A constraint is active when it has a nonzero reduced cost.

Organism	Model	Number of Active Constraints
<i>E. coli</i>	iJR904 [16]	2
<i>E. coli</i>	iAF1260b [17]	2
<i>L. lactis</i>	[18]	10
<i>L. plantarum</i>	[19]	4
<i>M. tuberculosis</i>	iNJ661 [20]	4
<i>S. cerevisiae</i>	iMM904 [21]	3

To visualize all possible product yields for a model, we use the enumeration of *Elementary Conversion Modes* (ECMs) [22,23]. The ECMs are the minimal stoichiometric relations that the cell can use to convert substrates into products and biomass, and all ECMs together comprise all its conversion capabilities. The set of ECMs therefore span all possible overall reactions from nutrients to products, and, importantly, each Elementary Conversion Mode has a corresponding Elementary Flux Mode that leads to the same conversion. Thus, if we can decide which ECMs are selected, we also know which EFMs are selected. It is beneficial to use ECMs rather than EFMs because they can be enumerated on much larger metabolic networks [22]. We will combine the enumerated ECMs with a graphical *cost vector formalism* [24] to obtain a figure that shows exactly the part of the metabolic capabilities of the model that is relevant for the specific FBA. Aided by this graphical formalism, it can be easily seen why a certain combination of minimal metabolic pathways is optimal.

We provide several examples that illustrate the benefit of analyzing why a certain FBA solution was selected. We use the cost vector formalism to show for a genome-scale model of *Escherichia coli* how much the FBA-solution changes when one gradually changes a constraint, as such providing insight in the robustness of the FBA-predictions (known as robustness analysis [25]). Genome-scale models of *Lactococcus lactis* and *Lactobacillus plantarum* are used to illustrate how our method facilitates a deeper understanding of FBA-results, even when many constraints are used to model nutrient-rich environments.

## 2. Results

### 2.1. Small Models Show that FBA Solutions Are Not Just Selected for Maximal Yield

In this section, we will use three small networks to identify the properties of metabolic pathways that determine whether the pathway will be used in the FBA-solution. A general Flux Balance Analysis problem can be written as

$$\max_v \{v_{\text{BM}} \mid Nv = 0, v_{\text{irrev}} \geq 0, v_{i_1} \leq C_{i_1}, \dots, v_{i_K} \leq C_{i_K}\}, \quad (1)$$

where  $N$  is the stoichiometry matrix that indicates which metabolites (rows) are used or produced in each reaction of the metabolic network (columns). It is multiplied by the flux vector,  $v$ , containing all reaction rates. Some of the reaction rates also appear separately:  $v_{i_1}$  denotes the first of  $K$  constrained uptake rates. Although in Equation (1) we have chosen to use only upper bounds, the results in this paper remain valid when these constraints are lower- or equality bounds. The objective reaction rate is denoted by  $v_{\text{BM}}$ : a virtual reaction combining all cell components in the right proportions for cell-synthesis. Finally,  $v_{\text{irrev}} \geq 0$  indicates that all irreversible reactions must have a non-negative rate.

Since uptake fluxes are often constrained, the ratio of product formation over substrate uptake plays an important role in FBA. This ratio is called the *product-substrate yield*,  $Y_S$ . For a flux vector,  $v$ , the biomass yield on substrate  $S$  is given by  $Y_S = v_{\text{BM}}/v_S$ . Note that this definition is completely general as to which product and which substrate is involved, we can thus have product-glucose yields, product-sulfur yields, and so on.

Any vector of reaction rates that satisfies the first two constraints of Equation (1) can be written as a combination of Elementary Flux Modes (EFMs) [3,26]. Since EFMs are minimal sets of reactions that can sustain a steady-state, these correspond to minimal metabolic strategies. To understand why a certain set of reactions is used in an FBA-solution, it is therefore sufficient to explain why the used EFMs were selected.

#### 2.1.1. FBA with One Constraint Selects the Maximal-Yield EFM

We here only consider *active* flux constraints, which means that  $v_{\text{BM}}$  could increase if the constraint was relaxed. This implies that  $v_S = C$  in the optimal solution. By using this and the definition of product-substrate yield, we can rewrite the FBA-optimization problem as

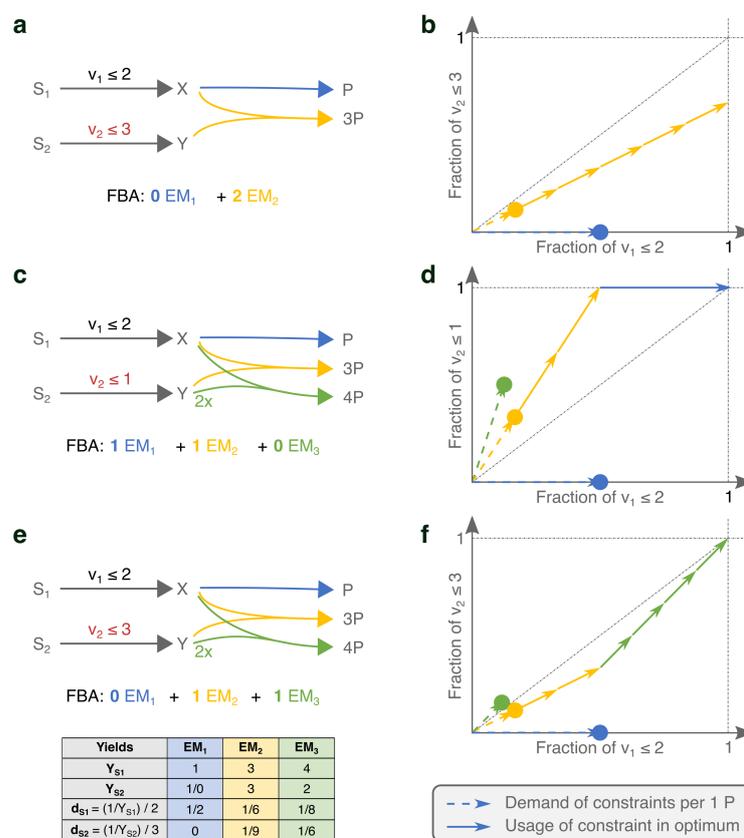
$$\begin{aligned} \max_v \{v_{\text{BM}} \mid Nv = 0, v_{\text{irrev}} \geq 0, v_S = C\} &= \max_v \{v_S Y_S \mid Nv = 0, v_{\text{irrev}} \geq 0, v_S = C\}, \\ &= \max_v \{C Y_S \mid Nv = 0, v_{\text{irrev}} \geq 0\}, \\ &= C \max_v \{Y_S \mid Nv = 0, v_{\text{irrev}} \geq 0\}. \end{aligned}$$

This confirms that the solution  $v$  is found by selecting the steady-state flux distribution with the highest yield. In Figure 1a, we show an example network with two EFMs and only one active constraint. The EFM that uses the yellow reaction has a three times higher product yield on the limiting substrate, and is therefore used in the FBA-solution.

It is thus easy to characterize solutions of Flux Balance Analyses with only one active constraint. One could, in principle, find the solution by enumerating the EFMs, and calculate the product-substrate yields for each of them. The EFM with maximal yield will give an optimal flux distribution. It may occur that several EFMs attain the maximal yield, and in that case, all combinations of these EFMs are optimal solutions, providing a mathematical basis for “flux variability” [27], see also [28].

In practice, the number of EFMs increases rapidly with network size, often rendering the above EFM-selection recipe infeasible [29]. Even though we are initially only interested in the product-substrate yields of the EFMs, EFM theory provides no way to speed up the enumeration by focusing only on that information. Elementary Conversion Modes (ECMs), however, do give that option [22,23]. The set of ECMs not only provides a full overview of all overall conversions from nutrients to products that a cell can catalyze; enumeration

with the recently introduced *ecmtool* also allows focusing only on the conversions between a selected set of nutrients and products [22]. This scales ECM enumeration to much larger networks than can be handled by EFM enumeration. Although ECMs are defined in terms of changes in external metabolite concentrations while FBA-constraints and -objectives are in terms of reaction rates, the important reaction rates can always be found by coupling the production of an external metabolite to each of these reactions. We will take a two-step approach in this paper: first, we compute the ECMs and use the conversion information to understand which ECMs constitute the FBA-solution. Then, we reconstruct a set of EFMs that gives rise to these overall conversions, because EFMs can be more easily interpreted as minimal metabolic strategies. Because these minimal strategies thus have both an EFM and an ECM associated to them, we will from now on use Elementary Mode, or EM, when the distinction does not matter.



**Figure 1.** Small networks show that the product-yields on limiting substrates in combination with the values of the upper bounds determine which EMs are selected in an FBA-solution. Panels (a,c,e) show networks of increasing complexity and the resulting FBA solutions. In (b,d,f), we show the respective visualizations of the costs of the various EMs with respect to the limited nutrient supplies. The dots indicate the fractions of the constrained supplies that the corresponding EM would use for producing one P. The dashed vectors help to emphasize that the costs of a combination of EMs can be visualized by using vector addition (as in [24]), but for visualization purposes, these dashed vectors will not be used in subsequent figures. The solid vectors show the usage by the various EFMs of the limited supplies in the FBA solution. The computation of the vector-positions is explained in the main text, and a calculation example is shown in (e). (a,b) show that the maximal yield EM is selected if there is only one active constraint. In (c,d) the uptake of  $S_2$  becomes limiting as well. Since the yellow and blue EMs are now selected, this illustrates that the product-yields on both substrates together determine which EMs are selected under two active constraints. We added the green EM to contradict that the maximal-yield EMs would always be selected. By relaxing the limit on the uptake of  $S_2$  in (e,f), we show that the precise values of the constraint bounds influence the selection of EFMs as well.

### 2.1.2. Under Two Constraints, EMs Are Selected on Two Different Yields

Can we now extend the above recipe—selecting the Elementary Mode with the maximal yield—to problems with several active constraints? Figure 1c shows that the answer is no. Although EM<sub>2</sub> produces more product per nutrient S<sub>1</sub>, the FBA solution is a combination of EM<sub>1</sub> and EM<sub>2</sub>. This is because the uptake of S<sub>2</sub> is now also actively constrained, suddenly making the product-yield per S<sub>2</sub> an important determining factor. We see that EM<sub>1</sub> has an infinite yield on S<sub>2</sub>, since it does not use it at all to produce P. So, while EM<sub>2</sub> has higher yield on S<sub>1</sub>, EM<sub>1</sub> is most efficient with S<sub>2</sub>. We can thus no longer simply say that FBA maximizes yield, since the FBA solution has a lower yield on both substrates than the individual EMs [30].

The FBA solution is also not a combination of the maximal-yield EMs with respect to the various constraints. For example, the addition of EM<sub>3</sub> in Figure 1c contradicts this intuition. The maximal-yield EMs are now EM<sub>3</sub> for S<sub>1</sub> and EM<sub>1</sub> for S<sub>2</sub>. EM<sub>2</sub> is second best in its usage of both substrates. Still, it is used in the FBA solution.

The examples in panels c and e of Figure 1 differ only in the value of one of the flux bounds ( $v_2 \leq 1$  versus  $v_2 \leq 3$ ), but still the selected EMs change. This is because we change ‘how limiting’ a constraint is by changing its bound, and with that the importance of the various substrate-yields change. This shows that it is not generally possible to have a rationale for the FBA-solution in terms of substrate yields only. Selection of the best combination of metabolic strategies should at least combine information about their substrate yields and the precise value of the bounds.

### 2.2. FBA Solutions Can Be Rationalized Using Cost Vectors

The above examples show that we can only understand which EMs are selected in an FBA solution by considering their product-yields on all limiting substrates, weighted by how limiting the various substrates are. It turns out that this is indeed all information that we need, and that this can be visualized with an intuitive graphical formalism. The derivation of this graphical formalism is based on a similar formalism that we developed for understanding resource-allocation optimization problems [24].

Consider a general EM with corresponding reaction rates  $v$ ; EMs are defined up to a positive constant, i.e., if we multiply  $v$  with an overall positive factor, it will still represent the same EM. To facilitate comparing the costs of making biomass between EMs, we start by rescaling all EMs such that one unit biomass is made:  $v \rightarrow v/v_{\text{BM}}$ . The resulting vector thus gives the reaction rates that are needed to produce one unit biomass via that EM. This can also be expressed in terms of the biomass-substrate yields:

$$v_{S_j} = \frac{v_{S_j}}{v_{\text{BM}}} v_{\text{BM}} = \frac{1}{Y_{S_j}^i} v_{\text{BM}} = \frac{1}{Y_{S_j}^i},$$

where we used  $v_{\text{BM}} = 1$  in the last equality. (Note that we could scale to any other objective flux.) The next step is to write this needed reaction rate as a fraction of the available rate. This gives a cost value,  $d_j^i$ , for EM  $i$  with respect to the  $j$ th constraint:

$$d_j^i = \frac{\frac{1}{Y_{S_j}^i}}{C_j} = \frac{1}{Y_{S_j}^i C_j}.$$

We can gather these costs in a so-called *cost vector* [24] for the  $i$ th EM:  $\mathbf{d}^i = [d_{S_1}^i \ \cdots \ d_{S_K}^i]^\top = [\frac{1}{Y_{S_1}^i C_1} \ \cdots \ \frac{1}{Y_{S_K}^i C_K}]^\top$ . These cost vectors can be compared in a  $K$ -dimensional cost vector plot (solid vectors in Figure 1b,d,f. The  $j$ th coordinate of the cost vectors thus shows the fraction of the  $j$ th constraint that is needed to produce one unit product with that EM. Cost vectors that are far from the origin thus indicate that the EM is inefficient in its substrate-usage and vice versa.

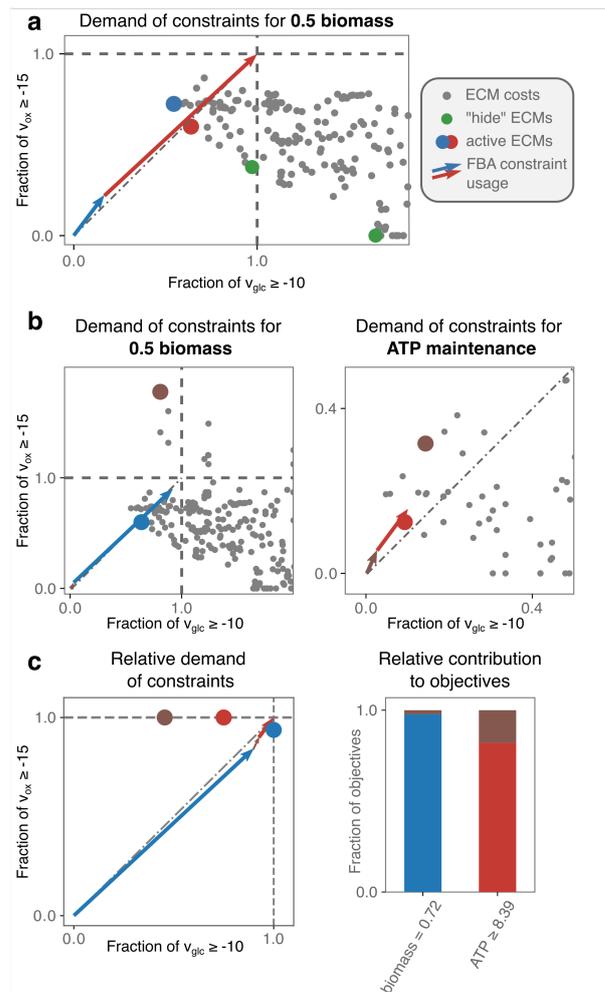
To see how the information in the cost vectors helps us understand the FBA solutions, we rephrase the FBA-optimization as: choose a combination of the rescaled EMs,  $\sum_{i=1}^M \lambda_i \mathbf{EFM}^i$ , such that the sum of multiples  $v_{\text{BM}} = \sum_{i=1}^M \lambda_i$  is maximized and all  $K$  constraints are satisfied:  $\sum_{i=1}^M \lambda_i d_j^i \leq 1$ . For the cost vectors, this is equivalent to: find the sum of cost vectors that uses the largest number of vectors and stays within the unit square (indicated by the gray dashed lines).

An important property of this type of optimization problem is that the solution is always a combination of a number of EFMs that does not exceed the number of active uptake constraints [24]. Recall that a constraint is called active if relaxing this constraint will lead to an increase in growth rate. We prove in Appendix A that this result remains valid if we use ECMs. We thus state that: *when  $K$  constraints are active, then there is an optimum with at most  $K$  EMs*. In the examples of Figure 1a,c,e, we, respectively, had 1, 2, and 2 active constraints, and indeed this was matched by the number of used EMs.

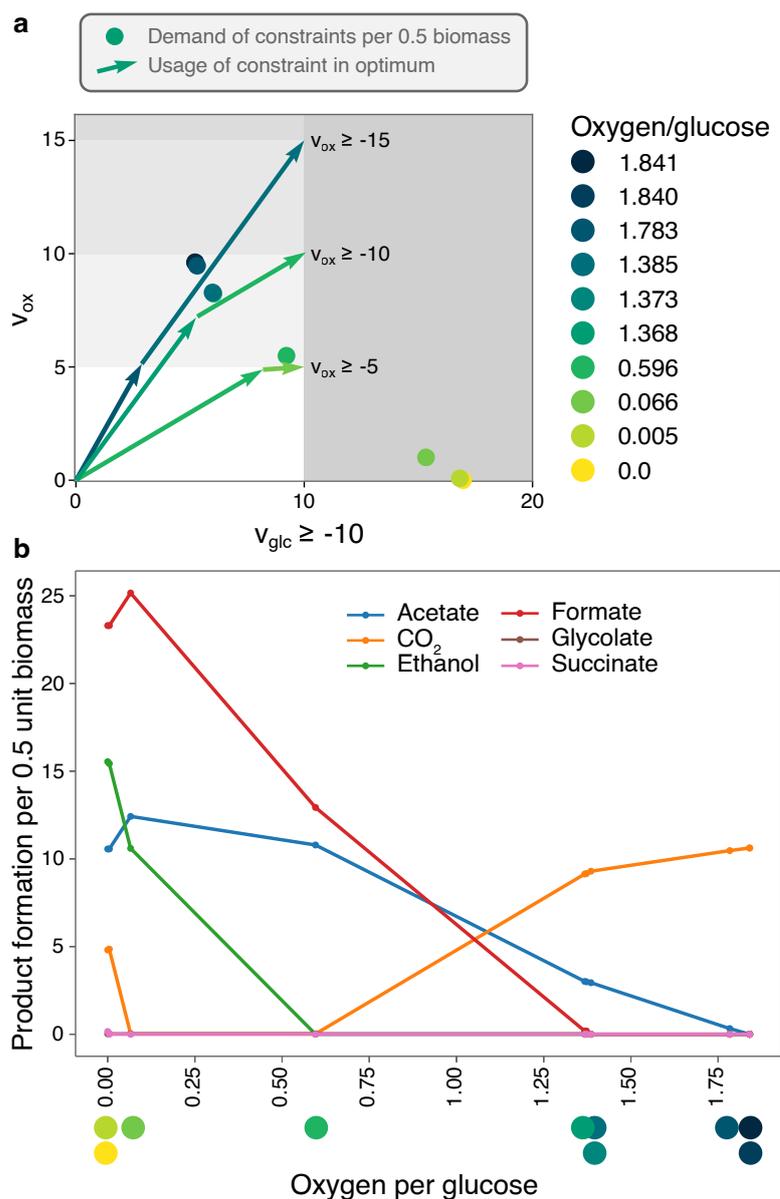
In problems with fewer than four constraints, we can even visually examine which combination of EMs is optimal [24], as we will describe below. However, even though this visual analysis was very useful in the case of resource-allocation because these problems often have few active constraints, FBA-problems may have many more constraints. Therefore, a visual optimization is often impossible, simply because we can only make 3-dimensional plots. Still, we will illustrate with various genome-scale examples that the cost vector formalism remains very useful in these multi-constraint cases: the cost vectors indicate the strengths and weaknesses of all EMs; a vector coordinate is low when the EM is very efficient with respect to that constraint.

In a case with two constraints (such as in Figures 1–3), the visual optimization relies on the knowledge that at most two EMs will be used in the optimum. Therefore, we only have to compare all candidate solutions using either a single EM or a pair of EMs. In [24], we have proven that we can rank all candidates by ordering them on the diagonal of the cost vector plot (see Figure A1 for an illustration of this scoring procedure). All EMs with a cost vector below the diagonal are scored by projecting their cost vector vertically to the diagonal, all EMs with an above-diagonal cost vector are scored by projecting the cost vector horizontally. Intuitively, this projection quantifies that the production rate of a single EM is limited by its weakest point, i.e., it is limited by the constraint that it uses most inefficiently (the largest coordinate of the cost vector). All pairs of EMs are scored by connecting their cost vectors by a line, and marking the intersection with the diagonal. As such, all candidates have been scored by a projection on the diagonal of the cost vector plot. The optimal candidate is the one that led to the lowest projection on the diagonal. This procedure immediately implies that all EMs that can be selected by FBA must have cost vectors that lie on the Pareto-front between efficient usage of the constraints. This can be clearly seen in Figures 2 and 3, as we will discuss shortly. When there are three active constraints, the visual optimization procedure is similar: each candidate combination of three EMs is projected to the diagonal by finding the intersection of the plane spanned by the three cost vectors and the diagonal. Eventually, the combination of EMs that gave rise to the lowest projection is the optimal one.

Although we here focused on the case where both the objectives and constraints involve only single reaction rates, cost vectors can be used whenever the objective and the constraints can be written as linear combinations of reaction rates. For example, when a combination of ATP production and biomass production is to be maximized ( $\max v_{\text{BM}} + v_{\text{ATP}}$ ), or when the summed uptake of glucose and lactose is bounded ( $v_{\text{glc}} + v_{\text{lac}} \leq C$ ). In Figure A2, we show that even when a ratio of fluxes is bounded ( $v_1/v_2 \leq C$ ) we can still use cost vectors by rewriting this as  $v_1 - Cv_2 \leq 0$ . The case in which a ratio of fluxes (or yield) is to be maximized ( $\max v_1/v_2$ ) under a set of different constraints ( $v_i \leq C$ ), is thoroughly analyzed in [30]. Often, such a maximal yield will be attained in a single EFM, and therefore we could find the optimal solution by just picking the EM with the highest yield. However, when the constraints prohibit the use of only this EM, the analysis needs to be extended. As far as we can see, this, however, does not fit inside our cost vector analysis.



**Figure 2.** A core model of *Escherichia coli* is used to introduce our analysis of FBA solutions. We analyze the FBA solutions of the *e\_coli\_core*-model [31] with a constraint on glucose- and oxygen-uptake in (a), and add an ATP-maintenance demand in (b,c). In (a), for each of the 689 ECMs, each dot represents the fraction of the constraints that the ECM takes up when it produces biomass at a rate of 0.5. We scale to 0.5 for visualization only; it does not change the analysis. The green dots represent the costs of the ECMs that are obtained using the `hide`-method of `ecmtool`. We explain in the main text (and in [22]) that these “hide”-ECMs are extreme with respect to the chosen constraints, and can therefore always form the optimal solution. The vectors show that the red and blue EMs constitute the FBA solution. This could have been found by the visual optimization procedure as well: when all pairs of cost dots are connected by a line, the line connecting the blue and red dot intersect the diagonal closest to the origin (see text and Figure A1 for an explanation). (b) An ATP-maintenance demand can be imposed as a constraint, but it is better to view it as setting a secondary objective: the demanded ATP should be produced as efficiently as possible. We therefore make an additional plot that indicates the costs of the EMs for meeting this demand. Not all EMs appear in both plots, e.g., an EM that does not produce ATP will not appear in the right plot because it has ‘infinite costs’. The selection of the optimal EMs can now be understood: the red EM produces ATP most efficiently; the rest of the available glucose- and oxygen-supply can be used by the blue EM to produce biomass. However, the red and blue EM cannot exactly use all of the supply: the remainder is used by the seemingly inefficient brown EM that produces both biomass and ATP. (c) With more than two constraints or secondary objectives, an alternative visualization will be better suited. We only visualize optimal EMs; their dots still indicate the relative usage of the constraints, but are no longer normalized with respect to an objective. The vectors show which EM uses which fractions of the constraints. In turn, the bar plot indicates which EM contributes to which objective.



**Figure 3.** The full flexibility of *E. coli* under glucose- and oxygen-limitation can be visualized with 10 EMs. We used the set of elementary conversions between glucose, oxygen, and biomass production that are possible in a genome-scale model of *E. coli* reported in [22]. For each ECM the demand of the constraints per 0.5 unit biomass is plotted in (a) and characterized with its oxygen per glucose requirement. (b) By reconstructing for each ECM a corresponding EFM (see Methods), we could reconstruct which products are secreted by each ECM. This information is given for all shown ECMs as a function of the oxygen/glucose requirement. The three different oxygen-bounds show that under increasing oxygen-limitation, the number of different fermentation products increases. When two neighboring dots are far from each other (a), there is a large regime of oxygen-limitation in which the corresponding EMs are used together. Some dots are so close together that the co-utilization of the EMs is highly unlikely.

### 2.3. Understanding FBA Solutions for Genome-Scale Models

We illustrate how Elementary Modes can be used to understand FBA solutions using several examples of increasing complexity.

### 2.3.1. A Core *E. coli*-Model as an Introduction

We show for a core model of *E. coli* how FBA-optimization can be fully understood with a two-dimensional plot when there are only two constraints Figure 2a. We plot the costs of all EMs with respect to the uptake constraints on glucose and oxygen. Dots that are placed closer to the origin represent EMs that are more substrate-efficient. Optimal FBA solutions can therefore always be constituted from the extreme EMs, i.e., the minimal set of EMs that spans the Pareto front that quantifies the trade-off between glucose- and oxygen-efficiency. These extreme EMs can be efficiently found with the `hide`-method of `ecmtool` (an explanation is given in Section 4). Because the `hide`-method facilitates ECM enumeration on larger networks and still captures all potentially optimal minimal strategies, we will often use this method on larger networks.

The full understanding of FBA solutions becomes harder, but not yet impossible, when a lower bound on some reaction rate is added. For example, in Figure 2b,c we add a lower bound on the production of ATP to model a non-growth related ATP maintenance. This constraint acts as a secondary objective: the ATP-demand has to be satisfied in the cheapest possible manner, so that the most substrate supply remains available for biomass synthesis. The ranking of EMs should now thus be done based on their efficiency in producing biomass, but also on their efficiency in producing ATP. We have therefore added a plot that shows the latter efficiencies.

It remains possible to rationalize the FBA solution. Two of the EMs (blue and red) are selected, respectively, for their substrate-efficiency in making biomass and ATP. However, combining only these two EMs would leave some of the oxygen supply unexploited. Therefore, the seemingly in-efficient brown EM can take over some of the ATP and biomass production while using relatively more oxygen than the red EM.

In larger models, the number of constraints often becomes larger than two, for example, because more nutrient uptakes are constrained. In addition, the number of secondary objectives may rise, for example, because uptake rates are fixed to their experimentally observed value, so that the uptake is forced. To facilitate the analysis in these more complicated cases, we introduce a simplified way of plotting the costs in Figure 2c. The left plot shows the costs of each EM with respect to all constraints; the right plot shows the contribution of each EM to the various objectives. These plots can be combined to understand why a certain EM was selected in the optimum.

### 2.3.2. Optimal Solutions under Glucose- and Oxygen-Limitation of a Genome-Scale *E. coli*-Model Are Determined by Few Minimal Strategies

Figure 3a shows the 10 EMs that together span all optimal solutions under glucose- and oxygen-limitation for the genome-scale *E. coli* model iJR904 [16]. Although ECM enumeration is often not possible on genome-scale networks, Clement et al. used the `hide`-method to only focus on conversions between glucose, oxygen, and biomass, and parallelized the enumeration to produce this set. We have taken the reported ECMs and reconstructed a corresponding Elementary Flux Mode for each of them (see Methods). This enables us to reconstruct the consumption and excretion of metabolites that were first hidden. Products that differ between the 10 EMs are shown in Figure 3b, as a function of their oxygen per glucose requirement.

Figure 3 shows that when oxygen-limitation becomes more stringent, the optimal EMs are increasingly selected for their oxygen-efficiency. Most optimal in terms of glucose yield is full respiration: since  $\text{CO}_2$  is the most oxidized product, the electrons from glucose all go to oxygen (Figure 3). This EM requires 1.84 oxygen per glucose. To save on oxygen per glucose, some of the  $\text{CO}_2$  needs to be converted into something less oxidized, and acetate is the next best candidate: compared to other fermentation pathways, acetate production yields one additional ATP after glycolysis, and the EM only needs 1.385 oxygen per glucose. Thus, this EM trades in a bit of the  $\text{CO}_2$ , produced in the citric-acid cycle, for acetate-production to maintain the redox balance. Since this also reduces the ATP yield on glucose, more glucose flux is required to produce one unit of objective flux. A

next oxygen-saving step is to replace  $\text{CO}_2$  by its more reduced form formate ( $\text{H}_2\text{CO}_2$ ). Again, the ATP yield drops, and more glucose is needed to produce biomass, but only 0.596 oxygen molecules are needed per glucose molecule. A further decrease in oxygen (to 0.066 per glucose) is possible by the production of the more reduced ethanol, and, finally, growth is possible under complete anaerobiosis by producing some succinate. Some  $\text{CO}_2$ -production is uncircumventable, as the citric-acid cycle needs to remain active for synthesis of biomass components. For the same reason, ethanol production is increased, as it oxidizes the additional electron-carrier NADH that is produced in the citric-acid cycle. Note that, even though the number of products seems to increase incrementally, we do see large shifts in product formation, most notably the replacement of almost all  $\text{CO}_2$ -production by formate-production. Incidentally, the production of formate requires the enzyme *pyruvate formate lyase*, which is unstable in the presence of oxygen. Therefore, this reaction can only be active in the last, anaerobic, EM in real *E. coli* cells, showing that what is stoichiometrically feasible, may not always be possible due to kinetic constraints.

Figure 3 also shows how robust a certain optimal solution is to changes in the nutrient-limitations. For example, at a fixed glucose-limitation of  $v_{\text{glc}} \geq -10$  and all oxygen-limitations between approximately  $V_{\text{ox}} \geq -7$  and  $V_{\text{ox}} \geq -12$ , the FBA solution will use a combination of the EM with oxygen/glucose requirement 1.368 and the EM with 0.596 oxygen/glucose requirement, producing  $\text{CO}_2$ , acetate, formate, and relatively small fractions of glycolate and ethanol. However, when oxygen-limitation is less stringent, there is only a very small region in which formate is produced, but not glycolate. This predicts that it is very unlikely to observe formate production without glycolate production. These regions of qualitatively different behavior are equivalent to the phenotype phase planes that were introduced in [32]. Our ECM-based method thus enables Phenotype Phase Plane (PhPP)-analysis without having to numerically scan the parameter space. This reduces potential numerical instabilities, and makes generalization to higher-dimensional ‘phenotype phase regions’ possible, where these could be defined as uptake rate regions that lead to the same set of active EMs.

### 2.3.3. *Lactococcus lactis* in Rich Medium: Elementary Modes-Based Analysis Can Rationalize Solutions to Many-Constraint Optimizations

Often, many constraints are used to model an organism’s metabolism, for example when the metabolic behavior of the organism is hard to predict from first principles. This could be because we have insufficient understanding of the constraints and the objective that determine this organism’s metabolic behavior. For example, not all organisms have been selected for maximal growth rates, and it is a much debated question which constraints shape metabolism the most [33]. In such cases, the goal of FBA can be shifted to deduce why the measured metabolic behavior may appear. Intuitively, the research question shifts from “What would the optimal metabolic behavior be given this set of constraints?” to “A certain metabolic behavior is observed, can we understand why, by analyzing the costs and benefits of this behavior?” In such cases, FBA can be used by fixing all measured rates, and then infer the rest of the metabolic state through growth rate optimization. This gives rise to a total of four constraint-types: (1) upper bounds on uptake rates (e.g., limited glucose-uptake), (2) lower bounds on production rates (e.g., an ATP-maintenance demand), (3) a lower bound on uptake rates (e.g., setting an uptake rate to its observed value), and (4) an upper bound on production rates. Constraint-types (2) and (3) can also be seen as secondary objectives: the FBA solution should satisfy these demands by using as little resources (the constraints) as possible. Our EM-based method can be used to understand the consequences of the various constraints, answering which constrained reaction is used by which metabolic strategy to meet which objective.

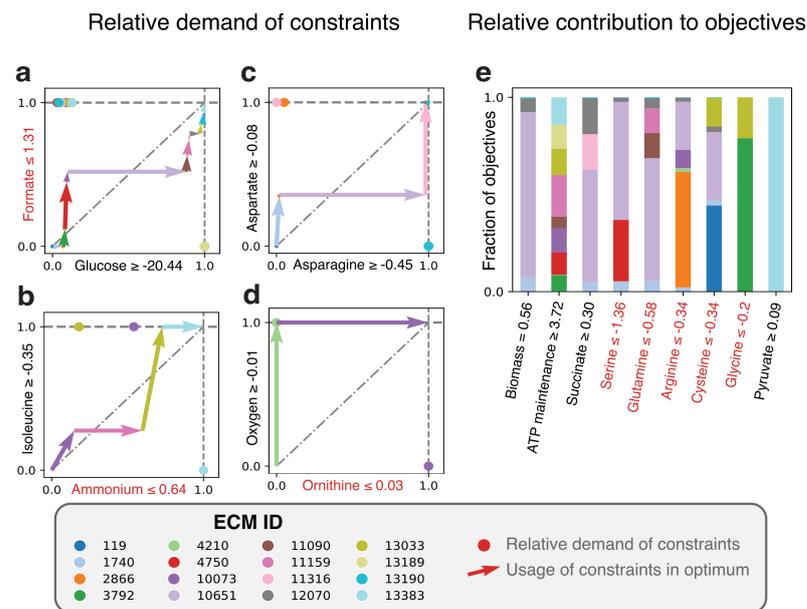
As an example, we use a genome-scale model of *Lactococcus lactis* [18]). This lactic acid bacterium has several amino acid auxotrophies [34,35], and is therefore grown in nutrient-rich media. Amino acids are not only used for protein synthesis, but are also catabolized to yield energy-equivalents [18,36]. It is therefore hard to explain from first principles why *L. lactis* limits its uptake of amino acids, even when these are available at unlimited

concentrations. Moreover, at high growth rates, lactic acid is used as a fermentation product, while at lower growth rates, it uses mixed acid fermentation, which produces more ATP per consumed carbon molecule [14,37]. It is unclear why the most carbon-efficient pathway is not always used. To still investigate the metabolic behavior, we fix certain reaction rates to their measured values, obtained from [37], in particular: amino acid uptake rates are limited, and the measured fermentation products are demanded to be excreted. Because the number of relevant metabolites is now very large (16), the `hide`-method is no longer enough to compute all ECMs on this genome-scale model. Therefore, we have selected the subnetwork containing only reactions that are active in the FBA-solution (see Methods for an explanation). This still yielded 14792 ECMs, of which 16 were used in the FBA-solution (as expected, because the number of ECMs is bounded by the number of constraints Appendix A [24]).

A cost-benefit analysis of all active EMs in the FBA-optimum is shown in Figure 4. Panels a–d indicate which EM takes up which part of the constraints, while panel e shows to which (secondary) objective each EM contributes. When first focusing on biomass formation, we note that the main contributor to biomass formation is EM 10651 (Figure 4e), which uses—as expected—the lactic acid fermentation pathway. EM 1740 is very similar to EM 10651, but differs in the asparagine and aspartate usage (see also the full ECMs of *L. lactis* in the Supplementary Materials). The other Elementary Modes that produce biomass do not produce lactate and differ in their relative amino acid usage, e.g., in the uptake of threonine. These differences may be of interest for industrial applications since threonine uptake has consequences for flavor formation (i.e., 2-Methylbutanal). Besides that the uptake of some components is limited, the experimentally observed uptake rates of some amino acids seem to be higher than needed, so that we must enforce these uptake rates to describe reality (red objectives in Figure 4e). For example, the forced uptake of arginine induces EM 10073 to produce ATP while ornithine is produced, which is shown to be an efficient ATP yielding conversion when glucose uptake is limiting [18]. While most of the other arginine consuming EMs use arginine in the same fraction, EM 2866 consumes a very large fraction of the arginine pool and only contributes a little to biomass formation. This gives rise to several questions: for example, would this EM be used to produce more ATP if ornithine production was not bounded? Would that not have been more efficient than currently consuming arginine with ECM 2866? If so, what inhibits *L. lactis* in doing so, or is this a solution that has just not been found by evolution? As this example shows, the Elementary Modes-based analysis facilitates the exploration of the metabolic strategies that give rise to the optimal solutions, which could lead to further computational and experimental investigations.

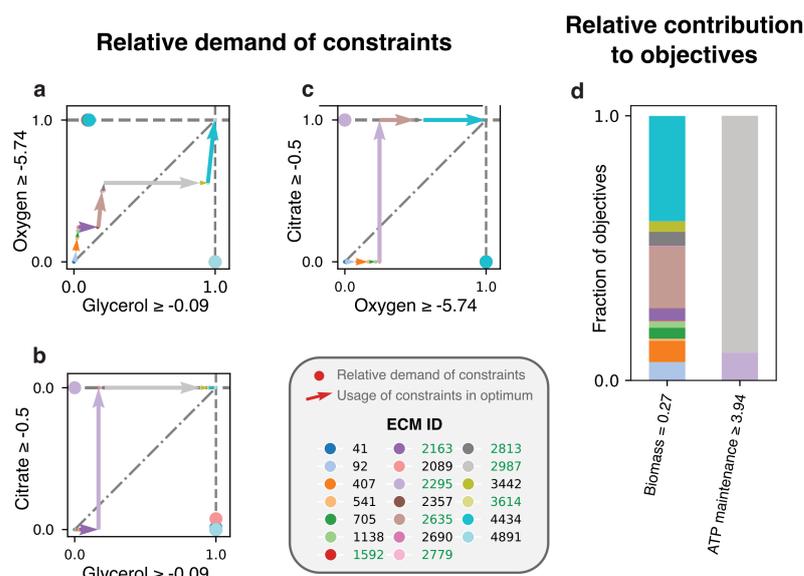
#### 2.3.4. *Lactobacillus plantarum*

EFM analysis of FBA solutions has been used before to understand the growth strategy of *Lactobacillus plantarum*. This analysis was done in a less systematic way, and could only be done for a model of central metabolism where ATP-production was optimized, because it was impossible to enumerate all EFMs for the genome-scale metabolic model. It is therefore interesting to see if we still find the same conclusions when we use our approach with a full-scale model and optimizing the biomass-production. In the original study, *L. plantarum* was grown on glycerol, which initially allowed only for very slow growth. After laboratory evolution, the growth rates and measured uptake fluxes turned out to be in agreement with the FBA predictions. It was surprising, however, that both in the experiments and in the FBA prediction, on glycerol, lactate was still produced rather than oxidized to acetate. EFM analysis showed that this can be rationalized by an oxygen uptake constraint: then it is optimal to ferment glycerol to lactate instead of acetate, because this can be done without using the limiting oxygen.



**Figure 4.** Elementary Modes-based analysis reveals the costs and benefits of all metabolic strategies that constitute the optimal solution of *Lactococcus lactis* in rich medium. ECMs were enumerated for the active network of a *L. lactis* genome-scale metabolic model [18] which was constructed based on measurements from [37]. (a–d) The costs of all the used EMs for the 8 different constraints. The pairing of the constraints in the plots is based on their scaled reduced costs, which quantifies how strongly the constraint limits the production of biomass. Glucose-limitation forms the strongest constraint, followed by formate, ammonium, etcetera. In principle, the relative costs between all pairs of constraints can be plotted. The constraints written in red are special: these force *L. lactis* to produce a certain minimal amount of product. (e) The contribution of each EM to the biomass objective and the secondary objectives. The secondary objectives are ordered based on their scaled reduced costs. Objectives written in red are special again: these do not demand the minimal production of something, but rather demand a certain minimal uptake rate.

Here, we redo this EM-based analysis for the full genome-scale network [19] with the objective to make biomass. Where the model of central metabolism only had constraints on glycerol, oxygen, and citrate, we now include constraints on amino acid uptakes as well. In addition, a non-growth related ATP maintenance is demanded, giving a total of 20 constraints. We found 5541 ECMs that were enumerated for the active subnetwork, and show the costs and benefits of the 20 ECMs that are active in the FBA-solution in Figure 5. We here opted to show only the costs for the constraints with largest scaled reduced costs: glycerol, oxygen, and citrate (see Figure A3 for an overview of all constraints). Figure 5d shows that only two EMs (2295 and 2987) are used to satisfy the ATP maintenance requirement. Although these consume all of the available citrate, and a large fraction of the available glycerol, they still produce lactate. Why is it not better to produce acetate and as such save some of the glycerol- and citrate-supply? The answer, confirming the conclusions drawn in [4], is that it is better to save all oxygen for biomass synthesis. The oxygen-limitation thus overrules the glycerol-limitation in this case, so that glycerol-yield is traded off for oxygen-yield. Interestingly, the original analysis still predicted that some ATP was produced with oxygen, while this is no longer predicted in our analysis. This is a consequence of using the genome-scale model: not only the usage of constraints of different EMs are important, but also their contributions to the various objectives.



**Figure 5.** *Lactobacillus plantarum* produces ATP at lower carbon-yields to save oxygen for biomass production. The ECMs in the active subnetwork of the genome-scale *L. plantarum* model [19] were enumerated based on measured fluxes adapted from [38], and we focused on the costs of EMs for the uptake of oxygen, glycerol, and citrate. (a–c) The uptake rates of oxygen, glycerol, and citrate by each active EM are shown. To fully show the relation between the costs for these three constraints, we made a different plot for each pair. (d) The contribution of the EMs to the production of non-growth related ATP and to biomass production is shown. An ECM ID is written in green if the ECM produces lactate. All citrate and a large fraction of the glycerol is used by ECMs 2295 and 2987 to meet the ATP-maintenance demand. Still, this production is done without the use of oxygen and while producing lactate, even though producing ATP with oxygen and by producing acetate would have a higher carbon-yield. It thus seems that all oxygen is needed for biomass synthesis, and therefore ATP is produced via a carbon-inefficient pathway.

### 3. Discussion

We have investigated the logic that underlies the optimality of FBA-solutions. It was already known that there is a simple rationale explaining the solutions of FBAs when only a single nutrient uptake rate is an active (limiting) constraint: FBA selects the Elementary Mode with the maximal yield on the growth-limiting nutrient. However, even though the modeling of many organisms requires multiple constraints, this rationale could not simply be extended to multi-constraint models [30]. We found that, in general, the EMs that constitute FBA solutions are selected for their production-yields with respect to all limiting uptake rates; the importance of these different yields is dependent on the values of the flux bounds. It turned out that these selection determinants could all be summarized in a cost vector formalism that allows for a visual examination of the strengths and weaknesses of all EMs, as such aiding the understanding of why a certain EM is selected by FBA or not.

We believe that this work forms a useful contribution to the field of genome-scale stoichiometric modeling. User-friendly toolboxes [39–43] and automated reconstruction algorithms [44–46] have made constraint-based metabolic models easy to run, but interpretation of the solutions is still very hard. Pitfalls like the high redundancy of optimal solutions [27,47], or high sensitivity of the solutions to user-defined constraints, remain key challenges. Hopefully, this may be overcome by providing a deeper understanding of the mathematical logic underlying the model outcomes.

In Figure 3, we showed, in one plot, all potential solutions to Flux Balance Analyses with glucose- and oxygen-limitation for a genome-scale model of *Escherichia coli*. This result extends a similar result by Carlson and Sreenc to genome-scale [48,49]. In this plot, we can for each combination of a glucose- and an oxygen-constraint, look up which EMs are optimal and why. Since only 10 EMs were needed for this plot, the optimal metabolism

of *E. coli* under glucose- and oxygen-limitation can be fully grasped by understanding 10 different minimal metabolic strategies. Additionally, the visual overview enables us to test how much a certain solution would change if the glucose- or oxygen-limitation would be relaxed or tightened. This may aid in examining which parts of the FBA-solution are robust predictions, and which parts are coincidental results of the specific flux bounds that were used.

In Figures 4 and 5, we have shown how solutions to multi-constraint models can be analyzed using cost vectors. In the case of *L. lactis*, for example, we investigated the amino acid usage by the various EMs. Comparing experimental observations with FBA-predictions showed that some amino acids are taken up at a much higher rate than needed. Our analysis gives an explanation: amino acids like arginine and serine can be used for ATP generation to meet the ATP maintenance demand. With *L. plantarum* we confirmed with a genome-scale model that it is optimal to anaerobically ferment glycerol and citrate to lactate to produce ATP, probably to preserve the limited oxygen supply for biomass formation. Studying the cost vectors of the same EMs for different pairs of constraints thus provides a glimpse into the strengths and weaknesses of the different EMs. This can be used to shape hypotheses answering why a certain metabolic strategy is selected, which can then be further analyzed. Such hypotheses may be of guidance to experimental perturbations that enhance or weaken the use of this strategy. Unfortunately, in this multi-constraint case, we can no longer use the visual optimization procedure that makes it possible to immediately see which EMs are used in the optimal solution under two constraints. This is simply due to our inability to make high-dimensional plots, and can thus not be overcome.

Although in our examples we have considered only prokaryotes, there is no reason that our analysis of FBA-solutions cannot be generalized to more complex cells or to cases where, for example, proteome constraints are considered instead of flux constraints. However, there may be cases in which no clear objective can be defined, for example, in the constituent cells of multi-cellular organisms, in which case our approach clearly does not work. We therefore only claim that our analysis works as long as we have a metabolic network reconstruction, a linear objective function, and linear constraints.

Our analysis of multi-constraint FBA solutions in terms of Elementary Modes was largely facilitated by two simplifying steps. First, Elementary Conversion Modes were enumerated instead of Elementary Flux Modes, which facilitates the enumeration of Elementary Modes on far larger networks [22,23]. This is mostly because there are generally much fewer ECMs than EFMs as many different EFMs map to the same overall conversion. In addition, ECM enumeration can focus on the conversions between a user-defined set of external metabolites, by ‘hiding’ information about the other metabolites. This can speed up the enumeration, and decrease the eventual number of ECMs by orders of magnitude (see [22] for a more elaborate discussion of this hide-method). We exploited this by only computing the Elementary Conversions from the limiting nutrients to biomass, since this gave exactly the necessary information to decide which of the ECMs would be used in the FBA solution. There seem to be some disadvantages to ECM-enumeration as well: (1) it does not return any information on which reactions are used, and (2) it disregards information on unconstrained metabolites that are taken up or secreted. However, this information can be retrieved easily for each ECM. We implemented a small Linear Program that finds an EFM for each enumerated ECM (described in Section 4). This shows which set of reactions could lead to the ECM, and what will happen to the first disregarded external metabolites.

Second, in the models of *L. lactis* and *L. plantarum*, we only computed the Elementary Conversion Modes for the subnetwork that was active in the FBA solution. For an FBA with only few constraints, this would have left us with a very small subnetwork, since the FBA solution would have used only a few EFMs [24]. Indeed, the active subnetwork on the core model of *E. coli* comprised only two EFMs. However, since we used the models of *L. lactis* and *L. plantarum* with around 20 constraints, there were also around 20 EFMs that thus gave rise to a much more versatile active network. Surprisingly, EFM enumeration

was not even possible on the resulting active subnetwork. The enumeration of ECMs while using the hide-method to ignore all unconstrained external metabolites still returned 14,792 ECMs for the *L. lactis* model and 5541 ECMs for the *L. plantarum* model. This once again illustrates the almost impermeable complexity of genome-scale models. Again, there seems to be a disadvantage to this simplification: when the active subnetwork is selected, we can no longer get the complete overview of why the selected metabolic state was better than all the alternatives. This is because there may be alternatives that are missing because they use a reaction that was not part of the active subnetwork. This is a challenge that we cannot solve at this moment, since the number of alternatives in genome-scale models with many constraints is just too large even for ECM enumeration.

The cost vectors that we used here were previously introduced for kinetic resource-allocation models. Although their translation to FBA-type problems was relatively straightforward, it still led to new insights that we did not obtain from their original application. This motivates their generalization to more types of models. We suspect that they can be used for any constraint-based model as long as Elementary Modes can be defined, and the costs of a certain mode scale proportionally to the usage of the mode. In the current case, this demand was satisfied: if the flux through an EFM increases by a factor, then the uptake of all limiting nutrients increases by the same factor. However, we have also tried to use cost vectors for Elementary Growth Modes [50], but the inherent non-linearity of self-fabrication made it impossible to use cost vectors. The cost vectors may thus not be readily applicable for self-fabrication models like Metabolism and Expression-models [51]. On the other hand, they may very well be applicable to the various extensions to FBA (for example described in [52]).

In the past decades, there have been several efforts to use an Elementary Modes-based perspective at FBA solutions. Poolman et al. [53] proposed a method that estimates how much each Elementary Flux Mode is used for a general flux distribution. Wiback et al. [54] extended this reconstruction by obtaining the lower and upper bounds of the possible usage of each Elementary Mode. However, both of these approaches made no attempt to explain *why* a certain set of EFMs was selected. There have also been various approaches where the yield values of all EFMs were plotted in a so-called *yield space* [55–57] to visualize the metabolic capabilities of a network. Our work may be seen as extending these approaches by focusing specifically on those yields that determine the FBA solution, and by quantifying the importance of the different yields by taking into account the flux bound value.

The usage of the cost vectors for FBA with only two constraints reproduces results that could have been obtained through Phenotypic Phase Plane (PhPP) analysis [32]. In this approach, the rates of two uptake reaction rates are varied and for each combination an FBA is run; the resulting biomass production rate is plotted in a 3D-plot. By considering the derivatives of the biomass rates with respect to the uptake reaction rates, the plane spanned by the variation of the uptake rates could be split into distinct regions with qualitatively different metabolic behavior. Similarly, in Figures 2 and 3, we could identify for which combinations of the flux bounds the FBA-solution would switch to a different pair of Elementary Modes. The Phenotypic Phase Planes are thus given by those regions in the space of constraint values for which the same Elementary Modes are selected. This makes it possible to generalize PhPP to FBAs with more than two constraints. Two additional advantages of our method is that it does not suffer from numerical instabilities when two EMs are very similar since the results of `ecmtool` are exact (working with fractions instead of floats), and that non-optimal Elementary Modes can also be plotted, so that it becomes clear why the optimal metabolic behavior is better than the alternatives.

## 4. Materials and Methods

### 4.1. The Use of ECMs Instead of EFMs

In this paper, we have consistently used ECMs in computations, even though we discuss EFMs during the interpretation. To understand why this is acceptable, we should

quickly recapitulate the definitions of the two Elementary Modes. To facilitate readability, we will not go into details here, but an elaborate explanation of the difference between ECMs and EFMs is given in [22]. We start with the steady-state flux cone:

$$\mathcal{F} = \{v \mid N_{\text{int}}v = 0, v_{\text{irrev}} \geq 0\}, \quad (2)$$

where  $N_{\text{int}}$  comprises the rows of the stoichiometry matrix  $N$  that correspond to internal metabolites. This cone thus comprises all steady-state flux vectors. The EFMs are defined as the *elementary vectors of the flux cone*, meaning that the set of EFMs is the minimal set such that each flux vector in  $\mathcal{F}$  can be written as a positive conformal sum of the EFMs. It is not necessary to go into the meaning of “conformal” here, but the interested reader is referred to [58].

Analogously, the Elementary Conversion Modes are defined as the *elementary vectors of the conversion cone*:

$$\mathcal{C} = \{\dot{c} = Nv \mid N_{\text{int}}v = 0, v_{\text{irrev}} \geq 0\} = N\mathcal{F}, \quad (3)$$

where the last equality indicates that the conversion cone is the image of the flux cone under the stoichiometry matrix. This means that  $\mathcal{C}$  contains all overall conversion between external metabolites that a cell can catalyze while in steady-state.

There are two facts about EFMs and ECMs that are important for this work. First, each ECM,  $\dot{c}$ , there exists an EFM,  $e$ , such that  $\dot{c} = Ne$ . In other words, for each Elementary Conversion, there is at least one Elementary Flux Mode that leads to that overall conversion. If we thus find an optimal minimal conversion, we know that we can find a corresponding minimal metabolic strategy. Second, all information about whether an EFM will be selected in the FBA solution is contained in the ECM it maps to.

The above two facts enable us to first use the ECMs to find out which overall conversions are optimal, and then reconstruct which EFMs lead to these conversions. It is true that many EFMs may lead to this one ECM, but this just reflects that there are often many optimal FBA-solutions [5,27].

As a final note, some EFMs do not map to a single ECM, but for example to a combinations of two ECMs. However, the yield-properties of such EFMs are interpolations of the two EFMs that lead to the ECMs, and therefore this intermediate EFM is never essential in constituting the FBA solutions. This is according to the intuition that everything that this EFM can do can also be done by at least one of the two other EFMs.

#### 4.2. Different Levels of Detail Can Be Used to Study FBA Solutions

In Figure 2, we have illustrated that we can use different sets of Elementary Conversion Modes to understand FBA solutions. We here describe the details of the three main sets of ECMs that we have used.

##### 4.2.1. ECMs between All External Metabolites for the Full Network

For small enough networks, we can use `ecmtool` to enumerate all minimal overall conversions from nutrients to products and biomass. For example, this gave 689 ECMs for the `e_coli_core`-model. Unfortunately, even though the number of ECMs scales better with network size than the number of EFMs, the enumeration of *all* ECMs on genome-scale networks is not yet feasible. The largest model for which all ECMs were enumerated until now is the `iIT341`-model of *Helicobacter pylori* comprising 485 metabolites and 554 reactions [22].

##### 4.2.2. ECMs between a Subset of External Metabolites for the Full Network

To facilitate our Elementary Mode based-analysis on genome-scale networks, we can use the `hide`-method that was implemented in `ecmtool`. This method gives all stoichiometric relations between a user-defined set of external metabolites, while the information about

the other metabolites is not shown (again we refer to [22] for an elaborate explanation). We no longer compute the elementary vectors of the conversion cone (Equation (3)), but of

$$C_{\text{shown}} = \{\dot{c}_{\text{shown}} = N_{\text{shown}}v \mid N_{\text{int}}v = 0, v_{\text{irrev}} \geq 0\}, \quad (4)$$

where  $N_{\text{shown}}$  is the matrix with only the rows of non-hidden external metabolites. Importantly, no information about the stoichiometric coupling between non-hidden metabolites is lost during this procedure. Therefore, if we hide all external metabolites of which the uptake or production is not constrained, we still get the full information that decides which overall conversions are selected in the FBA-solution. This information is sufficient to reconstruct an EFM that leads to this conversion, which also reveals the information about the hidden metabolites.

The hide-method generally leads to a smaller set of ECMs, and to faster enumeration. Logically, this effect increases with the number of external metabolites that are hidden. Therefore, this simplification step is most useful when we have few constraints, thus requiring few unhidden metabolites. For this reason, we could enumerate the ECMs on the full genome-scale iJR904-model of *E. coli* when we focused only on conversions between glucose, oxygen, and biomass (see Figure 3).

#### 4.2.3. ECMs between a Subset of External Metabolites for the Active Network

In case we want to understand FBA solutions for genome-scale networks that also have many constraints, we have to make another simplification. For the models of *L. lactis* and *L. plantarum*, we have chosen to focus only on the subnetwork of reactions that are used in the FBA solution. For this subnetwork, it then becomes possible to enumerate the ECMs and to analyze which of these ECMs are used in the optimum. In this case, we can no longer compare the optimal ECMs with all other ECMs, but the cost vector analysis remains useful to rationalize what each of the ECMs costs, and what it brings to the optimal combination. Besides this active subnetwork, one can think of other useful subnetworks that can be used. For example, we can select all reactions that can be active according to a Flux Variability Analysis [27], or one may use experimental information to decide which reactions may be active and should thus be taken into account.

#### 4.3. Outline of the Code

We here give a coarse description of the code that allows for the visualization of the costs with respect to the constraints.

- *Load model and set constraints.* Depending on the research question, one can leave the default constraints that are set in the model, or one can change them. This and the subsequent optimization steps are done with the `cbmpy`-package [39].
- *Perform FBA.* We initialize the model by performing an FBA. We use the minimization of the sum of absolute reaction rates as a secondary objective. This ensures that the number of Elementary Modes will indeed be bounded by the number of constraints.
- *Perform reduced cost analysis.* The reduced cost of a reaction can be defined as the derivative of the objective function with respect to the reaction value. The reduced costs thus capture what happens to the objective when the reaction value is increased. When a reduced cost is nonzero in the optimum, it must have hit a constraint, and these reactions are thus of interest to us. We then select only the inhomogeneous constraints, i.e., constraints with a bound that is nonzero. Now we are left with four categories of reactions: reactions can have a positive or a negative optimal rate, and the reduced cost can also be positive or negative. A positive rate with positive reduced cost indicates that the reaction is bounded from above, so this indicates a limited production. A positive rate with negative reduced cost indicates a forced production; negative rate with positive reduced cost indicates forced uptake, and negative rate with negative reduced cost means that this uptake is limited. We mark the reactions with forced production and uptake as secondary objectives, and the limited production and uptake reactions as constraints.

- *Coupling an external metabolite to each constraint.* Most constraints are on exchange reactions and therefore we can immediately use the exchange of the metabolite as a marker of the reaction flux when we compute ECMs. When an internal reaction rate is constrained, we use the tag-method of `ecmtool` to add an external metabolite that is produced in this reaction.
- *Optional: Simplifying the network for ECM-enumeration.* As explained above, we can use two simplifications: (1) we can choose to consider only the active network. This means that all reactions that are zero in the FBA-solution are deleted from the model. (2) We can choose to compute the ECMs only between the external metabolites related to a constraint. If we choose this option, we here find all `ECMtool`-indices of the metabolites that are to be hidden.
- *ECM enumeration.* For this step, we used `ecmtool` as a Python library (as opposed to running it via the command line), of which the code can be found at <https://github.com/SystemsBioinformatics/ecmtool> accessed on 14 April 2021. Our computations were done with a version available at 14 April 2021. It can also be installed via the Python package manager `pip`.
- *Compute costs for all ECMs with respect to each constraint normalized to each objective.* If we have  $n_{\text{ecm}}$  ECMs,  $n_{\text{obj}}$  objects, and  $n_{\text{cons}}$  constraints, we thus eventually get  $n_{\text{obj}}$  tables of dimensions  $n_{\text{ecm}}$  by  $n_{\text{cons}}$ . The rows of these tables give for each ECM the fraction of the various constraints that are used for the production of one unit objective.
- *Calculation of the activities of the ECMs in the FBA solution.* We first compute the overall conversion ( $\dot{c}_{\text{FBA}}$ ) that is induced by the FBA solution for the non-hidden metabolites. Then, we perform an LP that finds this conversion as a sum of ECMs:  $\dot{c}_{\text{FBA}} = \sum_i \lambda_i \mathbf{ECM}_i$ . We minimize the sum of the  $\lambda$ 's in this LP under the constraint that  $\lambda_i \geq 0$  for all  $i$ . In principle, we could also compute the  $\alpha$ -spectrum [47] for the various ECMs to see if they could contribute to the FBA solution, but we have not done that.
- *For each ECM, find a corresponding EFM.* Given an ECM  $\mathbf{ECM}_i$ , we can reconstruct an EFM by solving another LP: choose reaction rates  $v$  such that  $N_{\text{shown}} v = \mathbf{ECM}_i$ , with  $v_i \geq 0$ , and where  $v_i$  is minimized. The minimization of  $v_i$  ensures that we get a minimal solution, which will be an EFM.
- *Optional: reconstruct the full conversion if some metabolites were hidden.* Given the EFM,  $\mathbf{EFM}_i$ , we can now reconstruct the full conversion by multiplying with the full stoichiometry matrix:  $\dot{c} = N \mathbf{EFM}_i$ . As such, we get for each ECM a possible conversion between all external metabolites (i.e., including the first hidden metabolites).

#### 4.3.1. Code Availability

The full Python-code that was used to make all figures is freely available on Github at <https://github.com/SystemsBioinformatics/elementary-mode-based-FBA-rationale> (accessed on 14 April 2021). This repository contains example scripts that were used to generate the figures in this paper; these can be used to build upon.

## 5. Conclusions

We originally started this investigation with the question: “Does FBA maximize yield or rate?”. One of the first things that we have realized is that this is a badly-stated question for two reasons. First, we cannot speak of yield-maximization when we have more than one limiting nutrient: our toy networks (Figure 1) showed that introducing a second limiting nutrient immediately leads to different yields that are important. When discussing the ‘yield’ of a metabolic pathway, we must realize that we are merely discussing the ratio of two reaction rates, and we must thus specify which two reaction rates are considered. Second, FBA explicitly maximizes a production rate, so the question is rather trivial. A better question turned out to be: “Based on what properties does FBA select its optima?” That is what we have tried to answer here.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/metabo11050257/s1>. The Supplementary Materials Folder Results.zip contains the raw figures for the different models.

**Author Contributions:** Conceptualization, E.v.P.-K., D.H.d.G. and B.T.; formal analysis, D.H.d.G.; funding acquisition, B.T.; investigation, E.v.P.-K. and D.H.d.G.; methodology, E.v.P.-K. and D.H.d.G.; project administration, B.T.; resources, B.T.; software, E.v.P.-K. and D.H.d.G.; supervision, B.T.; validation, E.v.P.-K. and D.H.d.G.; visualization, E.v.P.-K. and D.H.d.G.; writing—original draft, D.H.d.G.; writing—review and editing, E.v.P.-K., D.H.d.G. and B.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by NWO VICI grant 865.14.005 and NWO grant ALWTF.2015.4.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** D.H.d.G. wants to thank Steffen Klamt for pointing out that yield was not always optimized in FBA, and thereby initiating this work.

**Conflicts of Interest:** The project is organized by and executed under the auspices of TiFN, a public-private partnership on pre-competitive research in food and nutrition. The authors have declared that no competing interests exist in the writing of this publication. Funding for this research was obtained from Friesland Campina, CSK Food Enrichment and the Netherlands Organisation for Scientific Research.

## Abbreviations

The following abbreviations are used in this manuscript:

FBA	Flux Balance Analysis
EFM	Elementary Flux Mode
ECM	Elementary Conversion Mode
EM	Elementary Mode

## Appendix A. An Upper Bound to the Number of Elementary Modes in the Optimum

We here provide a short mathematical proof that shows that the minimal number of Elementary Modes that is needed to constitute the FBA-solution is always bounded by the number of inhomogeneous constraints. This theorem is a small adaptation from the work in [24]. We will therefore refer to that work for the full proof and show only the work needed for the extension here.

We here use the term *Elementary Modes* to describe both EFMs and ECMs. To be able to couple reaction rates to an ECM, we assume that for each reaction of interest, there is an external metabolite that is uniquely produced in that reaction, such that the production of the metabolite exactly equals the reaction rate. Indeed, we describe in Section 4 that if such a metabolite does not exist in the original network, we can easily add one with the tag-method of `ecmtool`. Therefore, for each reaction rate that comes up in one of the constraints, we know that there is a corresponding reaction rate in each Elementary Mode.

As a prerequisite, we first define *cost vectors*, as they are most applicable to FBA-problems.

**Definition A1.** Consider an objective reaction rate,  $v_{obj}$ , and  $K$  linear constraints on reaction rates:

$$\sum_{j=1}^r w_j^{(k)} v_j \leq C^k, \quad \text{for } k \in \{1, \dots, K\}. \quad (\text{A1})$$

If Elementary Mode  $i$  corresponds to a set of reaction rates given by  $[v_1^{(i)}, \dots, v_{r-1}^{(i)}, v_{obj}^{(i)}]$ , then the cost vector of Elementary Mode  $i$  has entries

$$d_k^i = \frac{\sum_j w_j \frac{v_j^{(i)}}{v_{obj}^{(i)}}}{C^k}. \quad (\text{A2})$$

In addition, we define when a set of Elementary Modes is non-redundant.

**Definition A2.** A set of Elementary Modes is called non-redundant with respect to a set of constraints, if none of the corresponding cost vectors can be written as a positive linear combination of the others.

We are then ready to prove the theorem.

**Theorem A1.** Given a metabolic network with stoichiometry matrix  $N$ . Let  $v_{obj}$  be an objective flux which should be maximized at steady state, under  $K$  linear constraints on reaction rates:

$$\sum_{j=1}^r w_j^{(k)} v_j \leq C^k, \quad \text{for } k \in \{1, \dots, K\} \quad (\text{A3})$$

Then, at most,  $K$  non-redundant Elementary Modes are active in the optimal solutions.

**Proof.** We start by rescaling each Elementary Mode such that  $v_{obj} = 1$ . Given an arbitrary vector of reaction rates, we will in the following focus only on the objective rate and those rates that have a non-zero weight in one of the constraints of (A3). By definition of ECMs and EFMs, we can always write these combination of reaction rates as a conical sum of EFMs or ECMs

$$v = \sum_i \lambda_i \mathbf{EM}^{(i)}. \quad (\text{A4})$$

Because we have rescaled the Elementary Modes, we know that the rate of the objective reaction is given by  $\sum_i \lambda_i$ . In addition, we can express the constraints in terms of the weights  $\lambda_i$  and the cost vectors  $d_k^i$ :

$$C^k \geq \sum_j w_j^{(k)} v_j = \sum_j w_j^{(k)} \sum_i \lambda_i \mathbf{EM}_j^{(i)} = \sum_i \lambda_i \sum_j w_j^{(k)} \frac{v_j^{(i)}}{v_{obj}^{(i)}} = \sum_i \lambda_i C^k d_k^i, \quad (\text{A5})$$

so that we can rewrite all constraints as  $\sum_i \lambda_i d_k^i \leq 1$ . If we now collect the cost vectors in a cost matrix  $D = [d^1 \dots d^M]$ , then we can write the full optimization problem as

$$\max_{\lambda} \left\{ \sum_i \lambda_i \mid \lambda_i \geq 0, D\lambda \leq \mathbf{1} \right\}, \quad (\text{A6})$$

which is exactly the same optimization problem as described in Theorem 1 of [24]. We can thus use the result proven there that in the optimal solution at most  $K$  entries of  $\lambda$  will be non-zero, as long as the used EMs are non-redundant.  $\square$

Appendix B. Supplemental Figures

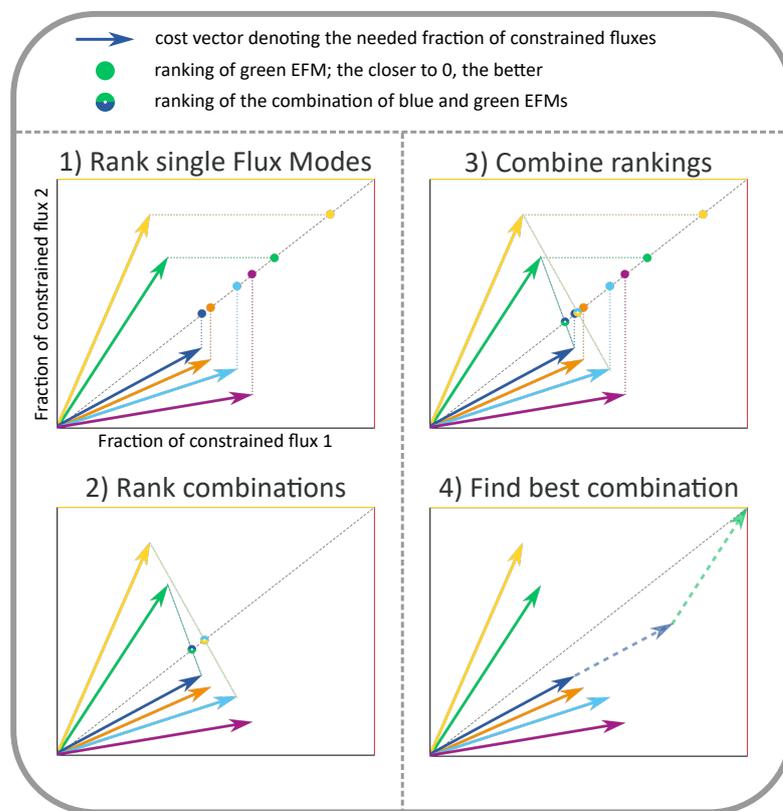


Figure A1. The optimal combination of EMs can be found by visual inspection of the cost vectors, as described in the main text.

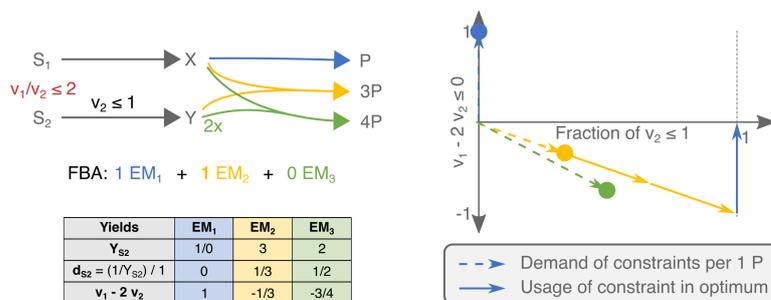
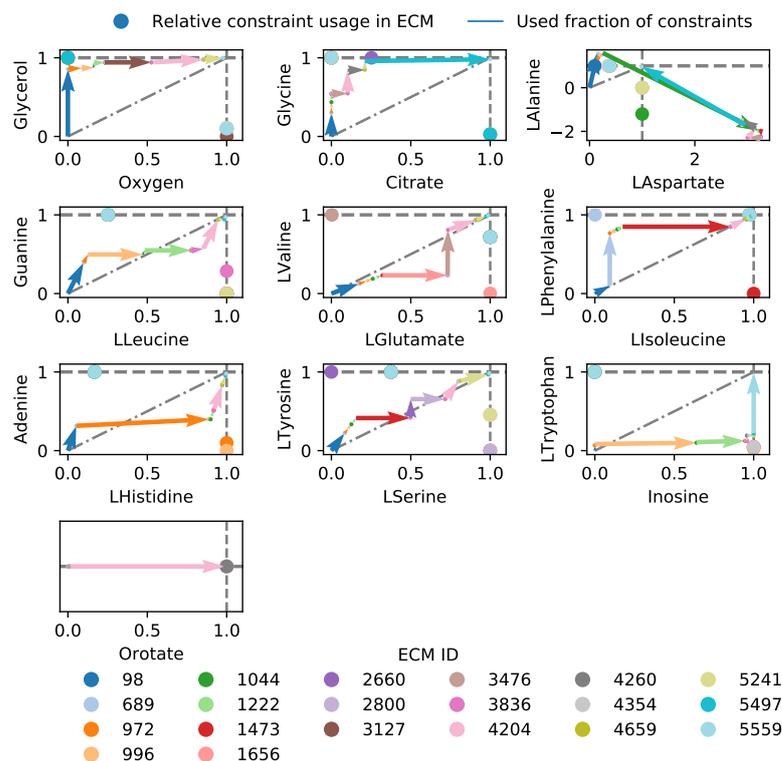


Figure A2. Our visual analysis of FBA solutions using cost vectors even works when constraints on ratios of reaction rates are used. The last toy network of Figure 1 is taken, and a constraint is added on the ratio of the uptake rates ( $v_1/v_2 \leq 2$ ). By rewriting this as  $v_1 - 2v_2 \leq 0$ , this becomes a linear constraint for which we can calculate the cost vectors: the second coordinate now captures the deviation from the imposed ratio that using this EM induces. A positive value indicates that the ratio of reaction rates used in this EM is too high, and vice versa. Where in the original optimization problem, the sum of cost vectors could not exceed the unit box, this sum may now not exceed the horizontal axis.



**Figure A3.** The costs of all the used EMs for the different constraints in *Lactobacillus plantarum*. The pairing of the constraints in the plots is based on their scaled reduced costs, which quantifies how strongly the constraint limits the production of biomass. Oxygen-limitation forms the strongest constraint, followed by glycerol, citrate, etc.

## References

- Orth, J.D.; Thiele, I.; Palsson, B.Ø.O. What is flux balance analysis? *Nat. Biotechnol.* **2010**, *28*, 245–248. [[CrossRef](#)]
- Raman, K.; Chandra, N. Flux Balance Analysis of Biological Systems: Applications and Challenges. *Brief. Bioinform.* **2009**, *10*, 435–449. [[CrossRef](#)]
- Schuster, S.; Hilgetag, C. On Elementary Flux Modes in biochemical reaction systems at steady state. *J. Biol. Syst.* **1994**, *2*, 165–182. [[CrossRef](#)]
- Teusink, B.; Wiersma, A.; Jacobs, L.; Notebaart, R.A.; Smid, E.J. Understanding the adaptive growth strategy of *Lactobacillus plantarum* by in silico optimisation. *PLoS Comput. Biol.* **2009**, *5*, 1–8. [[CrossRef](#)] [[PubMed](#)]
- Maarleveld, T.R.; Wortel, M.T.; Olivier, B.G.; Teusink, B.; Bruggeman, F.J. Interplay between constraints, objectives, and optimality for genome-scale stoichiometric models. *PLoS Comput. Biol.* **2015**, *11*, 1–21. [[CrossRef](#)] [[PubMed](#)]
- Ibarra, R.U.; Edwards, J.S.; Palsson, B.O. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* **2002**, *420*, 186–189. [[CrossRef](#)] [[PubMed](#)]
- Fong, S.S.; Joyce, A.R.; Palsson, B. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res.* **2005**, *15*, 1365–1372. [[CrossRef](#)] [[PubMed](#)]
- Schuster, S.; Pfeiffer, T.; Fell, D.A. Is maximization of molar yield in metabolic networks favoured by evolution? *J. Theor. Biol.* **2008**, *252*, 497–504. [[CrossRef](#)] [[PubMed](#)]
- Henry, C.S.; Broadbelt, L.J.; Hatzimanikatis, V. Thermodynamics-based metabolic flux analysis. *Biophys. J.* **2007**, *92*, 1792–1805. [[CrossRef](#)]
- Molenaar, D.; Berlo, R.V.; Ridder, D.D.; Teusink, B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol. Syst. Biol.* **2009**, *5*, 323. [[CrossRef](#)]
- Postma, E.; Verduyn, C.; Scheffers, W.A.; Van Dijken, J.P.; Scheffers, A.W.; Van Dijken, J.P. Enzymic analysis of the Crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **1989**, *55*, 468–477. [[CrossRef](#)] [[PubMed](#)]
- Wolfe, A.J. The acetate switch. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 12–50. [[CrossRef](#)] [[PubMed](#)]
- Warburg, O. On the origin of cancer cells. *Science* **1956**, *123*, 309–314. [[CrossRef](#)] [[PubMed](#)]
- Thomas, T.D.; Ellwood, D.C.; Longyear, V.M. Change from homo- to heterolactic fermentation by *Streptococcus lactis* resulting from glucose limitation in anaerobic chemostat cultures. *J. Bacteriol.* **1979**, *138*, 109–117. [[CrossRef](#)] [[PubMed](#)]

15. Folsom, J.P.; Carlson, R.P. Physiological, biomass elemental composition and proteomic analyses of *Escherichia coli* ammonium limited chemostat growth, and comparison with iron- and glucose-limited chemostat growth. *Microbiology* **2015**, *161*, 1659–1670. [CrossRef] [PubMed]
16. Reed, J.L.; Vo, T.D.; Schilling, C.H.; Palsson, B.O. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* **2003**, *4*. [CrossRef]
17. Feist, A.M.; Henry, C.S.; Reed, J.L.; Krummenacker, M.; Joyce, A.R.; Karp, P.D.; Broadbelt, L.J.; Hatzimanikatis, V.; Palsson, B. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **2007**, *3*. [CrossRef]
18. Chen, Y.; van Pelt-KleinJan, E.; van Olst, B.; Douwenga, S.; Boeren, S.; Bachmann, H.; Molenaar, D.; Nielsen, J.; Teusink, B. Proteome constraints reveal targets for improving microbial fitness in nutrient-rich environments. *Mol. Syst. Biol.* **2021**, *17*, e10093. [CrossRef]
19. Olivier, B.; Mendoza, S.N. SystemsBioinformaticsCbmpy-Metadraft: MetaDraft Is Now Available (Version v0.9.5). 2020. Available online: <https://zenodo.org/record/2398336#.YH-T9egzZPY> (accessed on 14 April 2021).
20. Jamshidi, N.; Palsson, B. Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. *BMC Syst. Biol.* **2007**, *1*. [CrossRef]
21. Mo, M.L.; Palsson, B.; Herrgård, M.J. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.* **2009**, *3*. [CrossRef]
22. Clement, T.J.; Baalhuis, E.B.; Teusink, B.; Bruggeman, F.J.; Planqué, R.; de Groot, D.H. Unlocking elementary conversion modes: ecmtool unveils all capabilities of metabolic networks. *Patterns* **2021**, *2*. [CrossRef] [PubMed]
23. Urbanczik, R.; Wagner, C. Functional stoichiometric analysis of metabolic networks. *Bioinformatics* **2005**, *21*, 4176–4180. [CrossRef] [PubMed]
24. De Groot, D.H.; van Boxtel, C.; Planqué, R.; Bruggeman, F.J.; Teusink, B. The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates. *PLoS Comput. Biol.* **2019**, *15*, 1–26. [CrossRef] [PubMed]
25. Edwards, J.S.; Palsson, B.O. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol. Prog.* **2000**, *16*, 927–939. [CrossRef] [PubMed]
26. Gagneur, J.; Klamt, S. Computation of elementary modes: A unifying framework and the new binary approach. *BMC Bioinform.* **2004**, *5*, 175. [CrossRef] [PubMed]
27. Mahadevan, R.; Schilling, C.H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **2003**, *5*, 264–276. doi/10.1016/j.ymben.2003.09.002. [CrossRef]
28. Kelk, S.M.; Olivier, B.G.; Stougie, L.; Bruggeman, F.J. Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. *Sci. Rep.* **2012**, *2*, 580. [CrossRef] [PubMed]
29. Klamt, S.; Stelling, J. Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.* **2002**, *29*, 233–236. [CrossRef] [PubMed]
30. Klamt, S.; Müller, S.; Regensburger, G.; Zanghellini, J. A mathematical framework for yield (vs. rate) optimization in constraint-based modeling and applications in metabolic engineering. *Metab. Eng.* **2018**, *47*, 153–169. [CrossRef]
31. Orth, J.D.; Palsson, B.Ø.; Fleming, R.M.T. Reconstruction and use of microbial metabolic networks: The core *Escherichia coli* metabolic model as an educational guide. *EcoSal Plus* **2010**, *4*. [CrossRef]
32. Schilling, C.H.; Edwards, J.S.; Letscher, D.; Palsson, B.O. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol. Bioeng.* **2000**, *71*, 286–306. [CrossRef]
33. de Groot, D.H.; Lischke, J.; Muolo, R.; Planqué, R.; Bruggeman, F.J.; Teusink, B. The common message of constraint-based optimization approaches: Overflow metabolism is caused by two growth-limiting constraints. *Cell. Mol. Life Sci.* **2020**, *77*, 441–453. [CrossRef]
34. Jensen, P.R.; Hammer, K. Minimal requirements for exponential growth of *Lactococcus lactis*. *Appl. Environ. Microbiol.* **1993**, *59*, 4363–4366. [CrossRef]
35. Teusink, B.; Molenaar, D. Systems biology of lactic acid bacteria: For food and thought. *Curr. Opin. Syst. Biol.* **2017**, *6*, 7–13. [CrossRef]
36. Fernández, M.; Zúñiga, M. Amino acid catabolic pathways of lactic acid bacteria. *Crit. Rev. Microbiol.* **2006**, *32*, 155–183. [CrossRef]
37. Goel, A.; Eckhardt, T.H.; Puri, P.; de Jong, A.; Branco dos Santos, F.; Giera, M.; Fusetti, F.; de Vos, W.M.; Kok, J.; Poolman, B.; et al. Protein costs do not explain evolution of metabolic strategies and regulation of ribosomal content: Does protein investment explain an anaerobic bacterial Crabtree effect? *Mol. Microbiol.* **2015**, *97*, 77–92. [CrossRef]
38. Teusink, B.; Wiersma, A.; Molenaar, D.; Francke, C.; De Vos, W.M.; Siezen, R.J.; Smid, E.J. Analysis of growth of *Lactobacillus plantarum* WCFS1 on a complex medium using a genome-scale metabolic model. *J. Biol. Chem.* **2006**, *281*, 40041–40048. [CrossRef] [PubMed]
39. Olivier, B.G. PySCeS CBMPy: Constraint Based Modelling in Python. 2011. Available online: <http://cbmpy.sourceforge.net/> (accessed on 14 April 2021).
40. Klamt, S.; Saez-Rodriguez, J.; Gilles, E.D. Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst. Biol.* **2007**, *1*, 2. [CrossRef] [PubMed]

41. Schellenberger, J.; Que, R.; Fleming, R.M.; Thiele, I.; Orth, J.D.; Feist, A.M.; Zielinski, D.C.; Bordbar, A.; Lewis, N.E.; Rahmanian, S.; et al. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nat. Protoc.* **2011**, *6*, 1290–1307. [[CrossRef](#)] [[PubMed](#)]
42. Rocha, I.; Maia, P.; Evangelista, P.; Vilaça, P.; Soares, S.; Pinto, J.P.; Nielsen, J.; Patil, K.R.; Ferreira, E.C.; Rocha, M. OptFlux: An open-source software platform for in silico metabolic engineering. *BMC Syst. Biol.* **2010**, *4*, 45. [[CrossRef](#)]
43. Hoops, S.; Sahle, S.; Gauges, R.; Lee, C.; Pahle, J.; Simus, N.; Singhal, M.; Xu, L.; Mendes, P.; Kummer, U. COPASI—a COMplex PATHway SIMulator. *Bioinformatics* **2006**, *22*, 3067–3074. [[CrossRef](#)]
44. Francke, C.; Siezen, R.J.; Teusink, B. Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol.* **2005**, *13*, 550–558. [[CrossRef](#)]
45. Thiele, I.; Palsson, B. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **2010**, *5*, 93–121. [[CrossRef](#)]
46. Mendoza, S.N.; Olivier, B.G.; Molenaar, D.; Teusink, B. A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biol.* **2019**, *20*, 158. [[CrossRef](#)] [[PubMed](#)]
47. Wiback, S.J.; Mahadevan, R.; Palsson, B. Reconstructing metabolic flux vectors from extreme pathways: Defining the  $\alpha$ -spectrum. *J. Theor. Biol.* **2003**, *224*, 313–324. [[CrossRef](#)]
48. Carlson, R.; Sreenc, F. Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: Creation of overall flux states. *Biotechnol. Bioeng.* **2004**, *86*, 149–162. [[CrossRef](#)] [[PubMed](#)]
49. Carlson, R.; Sreenc, F. Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: Identification of reactions. *Biotechnol. Bioeng.* **2004**, *85*, 1–19. [[CrossRef](#)] [[PubMed](#)]
50. De Groot, D.H.; Hulshof, J.; Teusink, B.; Bruggeman, F.J.; Planqué, R. Elementary Growth Modes provide a molecular description of cellular self-fabrication. *PLoS Comput. Biol.* **2020**, *16*, 1–33. [[CrossRef](#)]
51. Lloyd, C.J.; Ebrahim, A.; Yang, L.; King, Z.A.; Catoiu, E.; O'Brien, E.J.; Liu, J.K.; Palsson, B.O. COBRAme: A computational framework for genome-scale models of metabolism and gene expression. *PLoS Comput. Biol.* **2018**, *14*, e1006302. [[CrossRef](#)] [[PubMed](#)]
52. Gianchandani, E.P.; Chavali, A.K.; Papin, J.A. The application of flux balance analysis in systems biology. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2010**, *2*, 372–382. [[CrossRef](#)]
53. Poolman, M.; Venkatesh, K.; Pidcock, M.; Fell, D. A method for the determination of flux in elementary modes, and its application to *Lactobacillus rhamnosus*. *Biotechnol. Bioeng.* **2004**, *88*, 601–612. [[CrossRef](#)]
54. Wiback, S.J.; Mahadevan, R.; Palsson, B.Ø.O. Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: The *Escherichia coli* spectrum. *Biotechnol. Bioeng.* **2004**, *86*, 317–331. [[CrossRef](#)]
55. Klamt, S.; Mahadevan, R. On the feasibility of growth-coupled product synthesis in microbial strains. *Metab. Eng.* **2015**, *30*, 166–178. [[CrossRef](#)] [[PubMed](#)]
56. von Kamp, A.; Thiele, S.; Hädicke, O.; Klamt, S. Use of CellNetAnalyzer in biotechnology and metabolic engineering. *J. Biotechnol.* **2017**, *261*, 221–228. [[CrossRef](#)] [[PubMed](#)]
57. Nookaew, I.; Meechai, A.; Thammarongtham, C.; Laoteng, K.; Ruanglek, V.; Cheevadhanarak, S.; Nielsen, J.; Bhumiratana, S. Identification of flux regulation coefficients from elementary flux modes: A systems biology tool for analysis of metabolic networks. *Biotechnol. Bioeng.* **2007**, *97*, 1535–1549. [[CrossRef](#)] [[PubMed](#)]
58. Müller, S.; Regensburger, G. Elementary vectors and conformal sums in polyhedral geometry and their relevance for metabolic pathway analysis. *Front. Genet.* **2016**, *7*, 90. [[CrossRef](#)]